Computers in Biology and Medicine 64 (2015) 334–346

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/cbm

Computer-aided diagnosis system based on fuzzy logic for breast cancer categorization

Gisele Helena Barboni Miranda [1], Joaquim Cezar Felipe *,[1]

Department of Computing and Mathematics, Faculty of Philosophy, Sciences and Languages of Ribeirão Preto, University of São Paulo at Ribeirão Preto, Avenida Bandeirantes, 3900, Ribeirão Preto 14040-901, SP, Brazil

ARTICLE INFO

Article history:
Received 28 March 2014
Accepted 1 October 2014

Keywords:
Computer-aided diagnosis
Breast cancer
Fuzzy logic
Decision support
Healthcare informatics

ABSTRACT

Background: Fuzzy logic can help reduce the difficulties faced by computational systems to represent and simulate the reasoning and the style adopted by radiologists in the process of medical image analysis. The study described in this paper consists of a new method that applies fuzzy logic concepts to improve the representation of features related to image description in order to make it semantically more consistent. Specifically, we have developed a computer-aided diagnosis tool for automatic BI-RADS categorization of breast lesions. The user provides parameters such as contour, shape and density and the system gives a suggestion about the BI-RADS classification.
Methods: Initially, values of malignancy were defined for each image descriptor, according to the BI-RADS standard. When analyzing contour, for example, our method considers the matching of features and linguistic variables. Next, we created the fuzzy inference system. The generation of membership functions was carried out by the Fuzzy Omega algorithm, which is based on the statistical analysis of the dataset. This algorithm maps the distribution of different classes in a set.
Results: Images were analyzed by a group of physicians and the resulting evaluations were submitted to the Fuzzy Omega algorithm. The results were compared, achieving an accuracy of 76.67% for nodules and 83.34% for calcifications.
Conclusions: The fit of definitions and linguistic rules to numerical models provided by our method can lead to a tighter connection between the specialist and the computer system, yielding more effective and reliable results.

1. Introduction

Due to the increasing of the amount of digital medical images produced recently, the activities related to acquisition, management and availability of these images have required efforts of researchers and professionals involved in these processes to study and propose computational methods that provide effective support to daily tasks related to image-based diagnosis. Databases containing patient information as well as clinical data have become usual in healthcare environments. Machine learning and data representation techniques allow us to develop tools for computer-aided diagnosis (CAD) [1–3], which have been shown to improve the quality of the tasks performed by radiologists [4] and thereby the performance of radiological diagnosis [5].

Methods for CAD usually face a significant obstacle: the so-called *semantic gap*, which is characterized by the divergence between the results of automated methods and the results expected by users, due to the difficulties of these methods to represent real situations [6,7]. However, regardless of dealing with a specific context, a purely numeric representation of an image attribute is itself a semantic gap, since it is relatively distant from how the human analyst interprets the image. When analyzing the shape of a breast nodule, the diagnosis is generated in natural language, for example *slightly irregular contour*. In this case, it is very difficult to set a numeric descriptor for the contour that fits the user's semantics. So, we decided to use fuzzy concepts in this context. The concepts of fuzzy logic are designed to handle this kind of situation and can be used to simulate human reasoning in decision support systems [8]. Fuzzy sets handle the separation between classes, commonly limited by ranges of numeric values, by membership grades, replacing abrupt transitions. This approach is suitable to situations in which the boundary between the sets is not well delineated, e.g., the assessment performed by a human

* Corresponding author.
E-mail addresses: gimiranda@usp.br (G.H.B. Miranda), jfelipe@ffclrp.usp.br (J.C. Felipe).
[1] Tel.: +55 16 3602 4359; fax: +55 16 3633 9949.

http://dx.doi.org/10.1016/j.compbiomed.2014.10.006
0010-4825/

being. Fuzzy methods can approach objective knowledge (represented by numerical data) and subjective knowledge (represented by linguistic terms).

The Breast Imaging Reporting and Data System (BI-RADS) standard was developed in collaboration with members and committees of the American College of Radiology (ACR) to guide radiologists and physicians in the evaluation of breast nodules and calcifications [9]. This system aims to unify the terminology used in mammography exams, the categories of findings and their appropriate treatment [10]. The use of the BI-RADS lexicon is highly recommended in the decision-making process as it follows a natural language structure: "if the nodule is round, has well-defined margins and low density, it is more likely to be classified as benign". This kind of structure can be properly modeled through fuzzy sets. Fuzzy logic is also suitable to represent the divergence between radiologists. In clinical practice, the same image can be analyzed in different ways by each expert. Fuzzy inference models can encompass inter- and intra-observer variations.

In related works, some approaches have been proposed using fuzzy logic to assist the diagnosis of breast lesions. In Ref. [11] the authors present a method to distinguish lobulated contours from microlobulated ones. They created membership functions to represent features, such as the number of contour undulations and its size, and in Ref. [12] the authors implemented a fuzzy rule-based method aiming to characterize the shape of nodules in mammograms. In Ref. [13], three distinct regions in mammograms are compared: regions containing tumors; regions containing calcifications and regions containing no injuries. The authors proposed a fuzzy inference system based on the intensity of the pixels of these three regions, inferring the type of lesion (tumor or calcification) or the absence of lesion. In contrast, the methodology used in our project, in addition to treating attributes as fuzzy sets, also provides a fuzzy output, generating a quantitative response which is standardized by the BI-RADS system.

Fuzzy logic is also applied in Ref. [14], which presents an iterative segmentation algorithm to automatically identify regions of interest (ROIs) from nodules considering the breast tissue density and applying a fuzzy classifier for diagnosis. Concerning breast cancer treatment, Garibaldi et al. [15] present a method based on non-stationary fuzzy sets to model the need of recommendation of post-operative treatment considering clinical variables as attributes. This approach yielded better results when compared to standard fuzzy inference systems.

Another applications to support medical diagnosis can be found in Ref. [16] to help the analysis of malformations of the cerebral cortex. The input variables are based on the expert's knowledge. In Refs. [17,18], fuzzy cognitive maps are used to support diagnosis. In computer vision, fuzzy systems can also be used to describe intrinsic attributes of images [19,20]. Moreover, different studies have applied fuzzy-based techniques to medical applications [21–26].

This project aims to investigate the effectiveness of fuzzy representation of the evaluation parameters of breast lesions. For this purpose, we have developed a fuzzy system to provide an automated quantitative analysis of breast nodules and calcifications, based on image visual assessment parameters supplied by the users. This system is based on a new method proposed in Ref. [27] for designing of membership functions in fuzzy sets. This algorithm uses training sets to determine the data distribution for each class and establishes the design of the membership functions to be used by the fuzzy inference model. The system provides a second opinion to aid the radiologist (or other clinicians) to define the diagnosis and the patient's mammography classification. For this computerized analysis, we have used the main attributes related to mammographic findings in the reports of the nodules and calcifications. We have also used the BI-RADS evaluation categories and their corresponding prognosis. We have designed and implemented a CAD tool for generating BI-RADS diagnosis. We have evaluated our proposed methodology by the effectiveness in generating results more closer to the ones expected by users. Furthermore, the use of parameters and rules defined in terms of natural language should contribute to an effective adoption of CAD systems by the radiologists.

This paper is organized as follows: Section 2 presents the technical background related to fuzzy logic and the BI-RADS standard; Section 3 describes the proposed methodology for the representation of the BI-RADS attributes as fuzzy sets as well as their evaluation categories; Section 4 presents the experiments and the consequent results; Section 5 presents discussions and Section 6 presents conclusions.

2. Background

2.1. Fuzzy inference model

Fuzzy set theory distinguishes from the classical set theory since it concerns vague aspects related to information. This theory is less restrictive and can be more suitable to handle information given by humans. It extends the classical binary logic (0 *or* 1) to the continuous domain (0 *to* 1), through a gradual transition between pertinence and non-pertinence of an element within a set [8,28]. There are several fuzzy inference models. The most known is the Mamdani model, where the semantic rules use union and intersection operators (Max–Min) and both rule antecedents and consequents present fuzzy relations [29]. For each rule belonging to the database, the fuzzy Cartesian product (fuzzy intersection) is calculated using the components of the rule antecedent part generating an activation coefficient. When this coefficient is greater than zero, it is said that the rule was activated by the current input and will contribute in the calculation of the output provided by the system [30]. For the activated rules, a fuzzy union operation is applied, grouping the rules with the same output categories (or subsets) according to Eq. (1):

$$\mu(C) = \max_{k=1\cdots n}[\min(D_k,\ \mu C_i)] \tag{1}$$

where D_k is the activation coefficient of each rule that has the output C_i and $\mu(C)$ is the set of membership grades for each output class C_i. The set $\mu(C)$ can be used in the decision-making process in a qualitative assessment. For a quantitative result, or a scalar variable, it is necessary to perform the step of defuzzyfication, where $\mu(C)$ assumes a numerical value. The centroid method and the maximum average method are the most widely used [30].

2.2. BI-RADS

2.2.1. BI-RADS descriptors

The main attributes used for characterizing breast nodules and calcifications standardized by the BI-RADS system are described next [31].

2.2.1.1. Nodules: shape, contour and density. Shape can be classified as *round*, *oval*, *lobular* or *irregular*. Contour may assume the following values: *circumscribed*; *microlobulated*; *obscured*; *indistinct* and *spiculated*. The density of the nodules (compared to the underlying tissue) can be *high*, *isodense*, *low* or *fat containing*. Irregular shapes, spiculated contours and high densities usually implies a higher degree of malignancy for nodules. The combination of these three attributes characterizes the transition of a lesion from benign to malignant. Fig. 1 shows some examples of shape, contour and density in breast lesions.

2.2.1.2. Calcifications: type, distribution and quantity. Regarding the type, calcifications can be classified as *typically benign* (round skin, vascular, coarse etc.), *intermediate concern* and *higher probability of malignancy* (pleomorphic and linear). The distribution can be